

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-157091

(43)Date of publication of application : 31.05.2002

(51)Int. CI. G06F 3/06

G06F 12/16

(21)Application number : 2000-353010

(71)Applicant : HITACHI LTD

(22)Date of filing : 20.11.2000

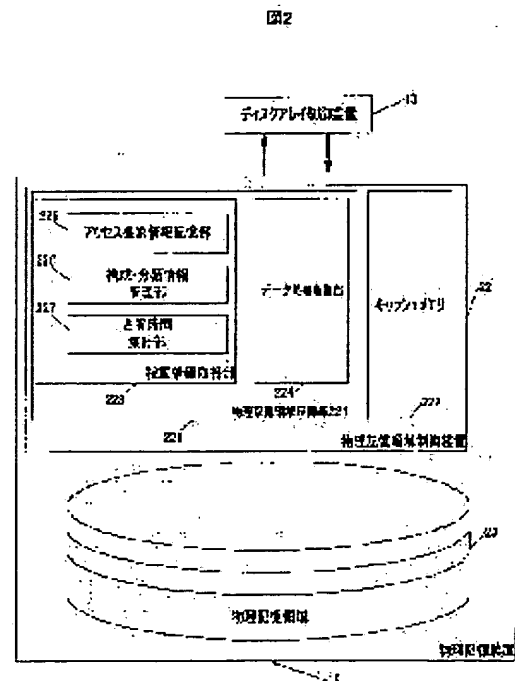
(72)Inventor : EGUCHI KENTETSU
MOGI KAZUHIKO
ARAKAWA TAKASHI
OEDA TAKASHI
ARAI HIROHARU

(54) STORAGE SUB-SYSTEM, AND MEMORY USED THEREFOR

(57)Abstract:

PROBLEM TO BE SOLVED: To obtain an occupied time of a logic storage area in a physical memory, and to obtain precise access occupied time information in every I/O to the physical memory.

SOLUTION: A physical storage area controller 22 on the individual physical memory 15 is provided with a table 225 for storing information about access requirement from a host computer, a table 227 for totalizing the occupied time as to access, a table 226 for control information for classifying constitution of a disk array, and a data processing control part 224 for obtaining constitution information and classification information of the logic storage area from a disk array controller 13, and for requesting the constitution information and the classification information of the logic storage area to the disk array controller, when necessary. The disk array controller 13 is provided with a means for transmitting the constitution information of the disk array at the present time to the physical storage area controller in response to the request from the physical storage area controller on the physical memory.



LEGAL STATUS

[Date of request for examination] 18.07.2003

[Date of sending the examiner's decision
of rejection]

[Kind of final disposal of application
other than the examiner's decision of
rejection or application converted
registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's
decision of rejection]

[Date of requesting appeal against
examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998, 2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2002-157091

(P2002-157091A)

(43) 公開日 平成14年5月31日 (2002.5.31)

(51) Int.Cl. ⁷	識別記号	F I	テームト* (参考)
G 0 6 F 3/06	3 0 2	G 0 6 F 3/06	3 0 2 A 5 B 0 1 8
	3 0 1		3 0 1 X 5 B 0 6 5
	3 0 4		3 0 4 N
	5 4 0		5 4 0
12/16	3 2 0	12/16	3 2 0 L
審査請求 未請求 請求項の数 5 O L (全 16 頁)			

(21) 出願番号 特願2000-353010(P2000-353010)

(22) 出願日 平成12年11月20日 (2000. 11. 20)

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 江口 賢哲

神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

(72) 発明者 茂木 和彦

神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

(74) 代理人 100093492

弁理士 鈴木 市郎 (外1名)

最終頁に続く

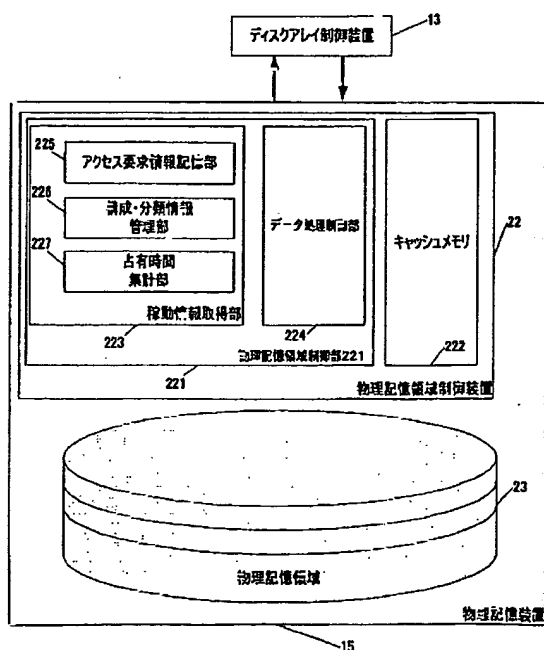
(54) 【発明の名称】 ストレージサブシステム及びそのシステムに使用する記憶装置

(57) 【要約】

【課題】 物理記憶装置において、論理記憶領域の占有時間を取得し、より精度の高い物理記憶装置へのI/O毎のアクセス占有時間情報を取得する。

【解決手段】 個々の物理記憶装置15上にある物理記憶領域制御装置22に、ホストからのアクセス要求に関する情報を記憶するテーブル225と、アクセスに関する占有時間を集計するテーブル227と、ディスクアレイの構成を分類する管理情報のテーブル226と、ディスクアレイ制御装置13から論理記憶領域の構成情報や分類情報を取得し、必要に応じて論理記憶領域の構成情報や分類情報をディスクアレイ制御装置にリクエストするデータ処理制御部224を備える。また、ディスクアレイ制御装置13に、物理記憶装置上にある物理記憶領域制御装置からのリクエストに応じて、現在のディスクアレイの構成情報を物理記憶領域制御装置に送信する手段を設ける。

図2



【特許請求の範囲】

【請求項1】 1または複数の計算機に接続され、複数の物理記憶装置と、これらの複数の物理記憶装置の使用状況情報を取得する手段と、前記計算機がリード／ライト対象とする論理記憶領域と前記物理記憶装置の物理記憶領域との対応付けを行う手段とを有するストレージサブシステムにおいて、前記複数の物理記憶装置のそれぞれは、物理記憶領域制御装置を備え、該物理記憶領域制御装置は、物理記憶領域の使用状況を取得する手段を有することを特徴とするストレージサブシステム。

【請求項2】 1または複数の計算機に接続され、複数の物理記憶装置と、これらの複数の物理記憶装置の使用状況情報を取得する手段と、前記計算機がリード／ライト対象とする論理記憶領域と前記物理記憶装置の物理記憶領域との対応付けを行う手段とを有するストレージサブシステムにおいて、前記複数の物理記憶装置の使用状況情報を取得する手段と、前記計算機がリード／ライト対象とする論理記憶領域と前記物理記憶装置の物理記憶領域との対応付けを行う手段とが、前述複数の物理記憶装置を制御する制御装置内に設けられ、前記制御装置は、さらに、物理記憶装置の論理記憶領域と物理記憶装置の物理記憶領域との対応付けを行った情報を前記複数のそれぞれの物理記憶装置に送信する手段を備え、前記複数の物理記憶装置のそれぞれは、物理記憶領域制御装置を備え、該物理記憶領域制御装置は、物理記憶領域の使用状況を取得する手段を有することを特徴とするストレージサブシステム。

【請求項3】 ストレージサブシステムを構成する物理記憶装置において、物理記憶領域制御装置を備え、該物理記憶領域制御装置は、物理記憶領域の使用状況を取得する手段を有することを特徴とする請求項1または2記載のストレージサブシステムに使用する物理記憶装置。

【請求項4】 前記物理記憶領域制御装置は、取得した物理記憶領域の使用状況情報を格納する手段をさらに備えることを特徴とする請求項3記載の物理記憶装置。

【請求項5】 前記物理記憶領域制御装置は、前記制御装置より受信する物理記憶領域の使用状況情報の取得要求に応じて、自物理記憶装置の物理記憶領域の使用状況情報を前記制御装置に送信する手段と、前記制御装置より受信する物理記憶装置の論理記憶領域と物理記憶領域とを対応付けした情報を格納する手段をさらに備えることを特徴とする請求項3または4記載の物理記憶装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、ストレージサブシステム及びそのシステムに使用する記憶装置に係り、特に、複数の記憶装置を有するストレージサブシステム及びそのシステムに使用する記憶装置に関する。

【0002】

【従来の技術】 コンピュータシステムに使用される高性能な二次記憶システムとして、ディスクアレイシステムが知られている。

【0003】 ディスクアレイシステムは、複数の物理記憶装置をアレイ状に配置し、各物理記憶装置にデータを分割して格納しておき、前記各物理記憶装置を並列に動作させて、前記各物理記憶装置に分割して格納されるデータのリード／ライトを高速に行うことを可能としたシステムである。

【0004】 ディスクアレイシステムに関する従来技術として、例えば、D.A.Patterson, G.Gibson, and R.H.Kats, "A Case for Redundant Arrays of Inexpensive Disks (ディスクアレイ)" (in Proc. ACM SIGMOD, pp. 109-116, June 1988) 等に記載された技術が知られている。この従来技術は、冗長性を付加したディスクアレイシステムに対し、その構成に応じてレベル1からレベル5の種別を与えておくというものである。また、これらの種別に、冗長性無しのディスクアレイシステムを加え、これをレベル0と呼ぶこともある。前述の各レベルは、冗長性等に応じて異なる構成として実現するため、コストや性能特性等が異なる。そして、ディスクアレイシステムを構築するにあたって複数のレベルのアレイ（物理記憶装置の組）を混在させることも多い。ここでは、冗長性を付加したディスクアレイの組をパリティグループと呼ぶ。また、物理記憶装置についても性能や容量等によりコストが異なり、ディスクアレイシステムを構築するにあたって最適なコストパフォーマンスを実現するために、やはり性能や容量の異なる複数種の物理記憶装置を用いることがある。

【0005】 ディスクアレイシステムに格納されるデータは、前述のような前記物理記憶装置に分散して配置される。このため、ディスクアレイシステムは、ディスクアレイシステムに接続されるホストコンピュータがアクセスする論理記憶領域と前記物理記憶装置の記憶領域を示す物理記憶領域の対応付け、すなわち、アドレス変換を行う必要がある。

【0006】 アドレス変換の処理を行うディスクアレイシステムに関する従来技術として、例えば、特開平9-274544号公報等に記載された技術が知られている。この従来技術は、ホストコンピュータからの論理記憶領域に対するI/Oアクセスについての情報を取得する手段と、論理記憶領域の物理記憶領域への対応付けを変更して物理的再配置を行う手段とにより、格納されたデータの最適配置を実現するというものであり、この公報には、論理記憶領域へのI/Oアクセス占有時間情報を、ディスクアレイ制御装置が取得する技術が開示されている。

【0007】 論理記憶領域へのI/Oアクセス占有時間情報は、ディスクアレイシステムの負荷分散を行うために、論理記憶領域の物理記憶領域への対応付けを変更し

て物理的再配置を行う際に、元になるデータとなるため重要である。

【0008】

【発明が解決しようとする課題】前述した公報に記載された従来技術は、論理記憶領域へのI/Oによる占有時間をディスクアレイ制御装置が取得するというものであるが、この従来技術に示された方法は、次に説明するような問題点を有している。

【0009】まず、ある論理記憶領域にデータの書き込み（ライト）が行われた場合を考える。この場合、当該論理記憶領域に対応する物理記憶領域にデータのライトが行われる。物理記憶領域は、物理記憶装置内に有り、物理記憶装置は、主に物理記憶領域制御部、データをキャッシュするキャッシュメモリ及び物理記憶領域により構成される。そして、データを物理記憶領域にライトする場合、物理記憶領域制御部がライトデータをキャッシュに書込んだ時点で書き込み終了の応答がディスクアレイ制御装置に通知される。このため、前述した従来技術は、実際にデータをライトするために物理記憶領域にアクセスした時間が判らないという問題点を有することになる。

【0010】次に、ある論理記憶領域にデータの読み込み（リード）が行われた場合を考える。この場合、当該論理記憶領域に対応する物理記憶領域にあるデータのリードが行われるが、実際には、物理記憶装置内のキャッシュメモリにそのリードデータがあった場合、物理記憶領域にはアクセスせず、キャッシュメモリにアクセスし、そのデータを返す。このため、前述した従来技術は、データリードのために実際に物理記憶領域にアクセスがあったかどうかを判別することができず、また、物理記憶領域にアクセスした正確な時間が判らないという問題点を有することになる。

【0011】また、A、B、C、Dと複数回にわたってある論理記憶領域にアクセスがあった場合を考える。そして、Aの応答をA'、Bの応答をB'等とし、アクセスAの時刻をA(t)、アクセスBの時刻をB(t)、応答A'の時刻をA'(t)等とする。ここで、アクセスAは、データリードで物理記憶装置にアクセスしてデータをリードし、アクセスBは、データリードでデータがキャッシュにあるため物理記憶装置にはアクセスしないでその応答B'があったものと仮定する。この場合、先にあったアクセスAの応答A'よりも後からきたアクセスBの応答B'の方が先になる。すなわち、 $A(t) < B(t)$ 、 $A'(t) > B'(t)$ となる。このとき、論理記憶領域へのI/Oによる占有時間をディスクアレイ制御装置において取得する従来技術は、どのI/Oが物理記憶装置内の物理記憶領域にどのくらいアクセスしたか、あるいは、物理記憶装置のキャッシュにヒットしたか否かをしることができないという問題点を生じる。

【0012】本発明の目的は、前述した従来技術の問題点を解決し、物理記憶装置で論理記憶領域の占有時間を取得することによって、ディスクアレイ制御装置のみでは取得不可能なシステム構成で各論理記憶領域の占有時間（実稼働時間）を取得できるようにしたストレージサブシステムを提供し、かつ、これに使用する記憶装置を提供することにある。

【0013】また、本発明の目的は、1つのI/O毎の物理記憶装置への影響を考慮することによって、前記ディスクアレイシステムの論理記憶領域の利用率の解析や利用率予測の誤差を小さくすることができ、より最適な性能チューニングを行うために、より精度の高い物理記憶装置へのI/O毎のアクセス占有時間情報を取得することができるストレージサブシステムを提供し、かつ、これに使用する記憶装置を提供することにある。

【0014】

【課題を解決するための手段】本発明によれば前記目的は、1または複数の計算機に接続され、複数の物理記憶装置と、これらの複数の物理記憶装置の使用状況情報を取得する手段と、前記計算機がリード/ライト対象とする論理記憶領域と前記物理記憶装置の物理記憶領域との対応付けを行う手段とを有するストレージサブシステムにおいて、前記複数の物理記憶装置の使用状況情報を取得する手段と、前記計算機がリード/ライト対象とする論理記憶領域と前記物理記憶装置の物理記憶領域との対応付けを行う手段とが、前述複数の物理記憶装置を制御する制御装置内に設けられ、前記制御装置が、さらに、物理記憶装置の論理記憶領域と物理記憶装置の物理記憶領域との対応付けを行った情報を前記複数のそれぞれの物理記憶装置に送信する手段を備え、前記複数の物理記憶装置のそれぞれが、物理記憶領域制御装置を備え、該物理記憶領域制御装置が、物理記憶領域の使用状況を取得する手段を有することにより達成される。

【0015】また、前記目的は、ストレージサブシステムを構成する物理記憶装置において、物理記憶領域制御装置を備え、該物理記憶領域制御装置が、物理記憶領域の使用状況を取得する手段、取得した物理記憶領域の使用状況情報を格納する手段、前記制御装置より受信する物理記憶領域の使用状況情報の取得要求に応じて、自物理記憶装置の物理記憶領域の使用状況情報を前記制御装置に送信する手段、及び、前記制御装置より受信する物理記憶装置の論理記憶領域と物理記憶領域とを対応付けした情報を格納する手段を備えることにより達成される。

【0016】

【発明の実施の形態】以下、本発明によるストレージサブシステム及びそのシステムに使用する記憶装置の実施形態を図面により詳細に説明する。

【0017】図1は本発明によるストレージサブシステムを備えた計算機システムの構成を示すブロック図、図

2は物理記憶装置の構成を示すブロック図である。図1、図2において、10はホスト、12はストレージサブシステム、13はディスクアレイ制御装置、14はディスクアレイ制御情報、15は物理記憶装置、16はディスクアレイ、17は制御端末、18はI/Oバス、19はネットワーク、22は物理記憶領域制御装置、23は物理記憶領域、130はリード/ライト処理部、131は使用状況情報取得処理部、132は再配置判断処理部、133は再配置実行処理部、141は論理/物理対応情報、142はクラス構成情報、143はクラス属性情報、144は論理領域使用状況情報、145は物理領域使用状況情報、146は再配置判断対象期間情報、147は再配置実行時刻情報、148は未使用領域情報、149は再配置情報、14Aは記憶装置占有時間情報、221は物理記憶領域制御部、222はキャッシュメモリ、223は稼働情報取得部、224はデータ処理制御部、225はアクセス要求情報記憶部、226は構成・分類情報管理部、227は占有時間集計部である。

【0018】図1に示す計算機システムは、上位の計算機である1または複数のホスト10、ストレージサブシステム12、制御端末17から構成される。ホスト10は、ストレージサブシステム12にI/Oバス18で接続され、ストレージサブシステム12に対してデータのリードやライト処理のためのI/Oを発行する。このI/Oを行う際、ホスト10は、ストレージサブシステム12の論理的な記憶領域を指定する。すなわち、ホスト10は、ストレージサブシステム内のデータに対して、通常論理的な記憶領域のアドレスによりアクセスを行う。また、I/Oバス18は、例えば、ESCON、SCSI、ファイバチャネル等により構成される。

【0019】ストレージサブシステム12は、ディスクアレイ制御装置13及び複数の物理記憶装置15から構成される。ディスクアレイ制御装置13は、リード/ライト処理部130と、使用状況情報取得処理部131と、再配置判断処理部132と、再配置実行処理部133とを備え、これらの処理部が、リード/ライト処理、使用状況情報取得処理、再配置判断処理、再配置実行処理等の処理を行う。また、ストレージサブシステム12のディスクアレイ制御装置は、論理記憶領域/物理記憶領域対応情報141、クラス構成情報142、クラス属性情報143等のディスクアレイ構成情報1400と、論理領域使用状況情報144、物理領域使用状況情報145等の記憶装置占有時間情報14Aと、再配置判断対象期間情報146と、再配置実行時刻情報147と、未使用領域情報148と、再配置情報149等を保持している。なお、前述したディスクアレイ構成情報14には、前述した情報の他、パリティグループ情報やRAIDレベル情報等が含まれてもよい。

【0020】また、ホスト10、ディスクアレイ制御装置13及び制御端末17は、相互にネットワーク19に

より接続されている。ネットワーク19は、例えば、イーサネット（登録商標）、FDDI、ファイバチャネル等により構成されてよい。制御端末17は、通常、ストレージサブシステム12の保守・管理等を行うために使用される。

【0021】また、ホスト10、ディスクアレイ制御装置13及び制御端末17には、それぞれでの処理を行うためのメモリ、CPU等の計算機において必ず存在する構成要素をそれぞれ存在するが、本発明の実施形態の説明においては重要でないため、ここでは明記しない。

【0022】前述のストレージサブシステム12内に設けられる複数の物理記憶装置15は、物理記憶装置の性能毎にクラス分けされて、クラス毎にディスクアレイ16を構成している。また、ここでは、明示的に示していないが、複数の物理記憶装置を使用して、パリティグループが構成されている。そして、物理記憶装置15のそれぞれは、図2に示すように、物理記憶領域23とこの物理記憶領域23を制御する物理記憶領域制御装置22とにより構成され、物理記憶領域23には、様々なデータが格納されている。

【0023】また、前述したように、ホスト10からは物理記憶領域23のアドレスは直接見えてはならず、ホスト10は、複数の物理記憶領域23上にある複数の論理的な記憶領域上にあるデータにアクセスを行う。すなわち、ホスト10は、ストレージサブシステム12内の各物理記憶装置15の記憶領域にあるデータに対して、論理記憶領域を指定してアクセスを行う。

【0024】ディスクアレイ制御装置13は、複数の物理記憶装置15と接続されており、複数の物理記憶装置15を制御したり、前記ホスト10から発せられたリードやライト処理命令I/Oを、指定のデータが存在する論理記憶領域のアドレスとその論理記憶領域のアドレスがある物理記憶領域のアドレスとを対応させて、適当な物理記憶装置15にデータI/Oを送信し、ライト処理であれば、ホスト10から送信されてくるデータを物理記憶装置15に送信し、リード処理であれば物理記憶装置15から送信されてくるデータを受信してホスト10に送信する等の処理を行っている。

【0025】物理記憶装置15内に備えられる物理記憶領域制御装置22は、物理記憶領域制御部221とキャッシュメモリ222とにより構成されている。キャッシュメモリ222は、物理記憶領域23に比べデータのリード/ライトの処理の速度が速い。そして、キャッシュメモリ222は、ディスクアレイ制御装置13から送信されてくるリードまたはライト命令に関するデータに関して次のように使用される。すなわち、ライト処理の場合、ディスクアレイ制御装置13から送信されてくるライトデータが物理記憶領域23に書き込まれる際に、データは、キャッシュメモリ222にも書き込まれる。また、リード処理の場合、物理記憶領域23からデータ読

み出される際に、読み出されたデータは、キャッシュメモリ222に書き込まれ、あるいは、以前のリード処理によって同一のデータがキャッシュメモリに有り、そのデータに対してディスクアレ制御装置13からリード命令として物理記憶装置にきた場合に、そのデータを物理記憶領域23から読み込まず、キャッシュメモリ222から読み込む。これにより、物理記憶装置15の処理性能を上げることができる。

【0026】物理記憶領域制御部221は、主に移動情報取得部223とデータ処理制御部224とを備えて構成されている。データ処理制御部224は、ディスクアレ制御装置13から送信されてくるデータのリードまたはライト命令を受信する。そして、データ処理制御部224は、受信した命令がリード命令であった場合、キャッシュメモリ222にアクセスし、そのリードデータがキャッシュメモリ222に存在すれば、キャッシュメモリ222からそのリードデータを読み出し、キャッシュメモリ222にそのデータがなければ、物理記憶領域23にアクセスしてそのリードデータを読み出して、ディスクアレ制御装置13にデータを送信する。また、データ処理制御部224は、受信した命令がライト処理命令であった場合、ディスクアレ制御装置13から送信されてくるデータをキャッシュメモリ222に書き込むと同時に、またはその後で、そのデータを物理記憶領域23に書き込む。ライトデータは、キャッシュメモリ222に書き込まず、直接物理記憶領域23に書き込んでもよい。

【0027】移動情報取得部223は、アクセス要求情報記憶部225、構成・分類情報管理部226、占有時間集計部227等により構成される。前述のデータ処理制御部224は、I/O処理によって指定されたデータのある論理記憶領域や、そのデータが存在する物理記憶領域23、あるいは、キャッシュメモリ222にアクセスしたときに、そのアクセスの時間情報をI/O処理の処理種別毎（ランダムアクセスかシーケンシャルアクセスか等）に分類して占有時間集計部227に記録する。また、データ処理制御部224は、ディスクアレ制御装置13よりディスクアレ内の論理記憶領域のアドレスと物理記憶領域23のアドレスとの対応情報や物理記憶装置15の性能等の情報を受信し、構成・分類情報管理部226に記録する。さらに、データ処理制御部224は、複数のI/O処理を受け付けることが可能なように、ディスクアレ制御装置13等から送信されてくるデータのリードまたはライト命令データ等を受信し、アクセス要求情報記憶部225にその命令データを記録する。

【0028】物理記憶装置15は、前述したような構成を有することにより、物理記憶装置15内でI/O処理による論理記憶領域、物理記憶領域23、あるいは、キャッシュメモリ222にアクセスしたときのアクセス時

の時間情報をI/O処理の処理種別毎（ランダムアクセスかシーケンシャルアクセスか等）に分類して占有時間集計部227に記録することが可能となり、I/O処理によって、物理記憶領域にどのくらいの時間アクセスしたか、あるいは、物理記憶装置15のキャッシュメモリ222にヒットしたかを分類して、その占有時間の集計を行うことが可能となる。

【0029】図3はストレージサブシステムが起動されたときのディスクアレ制御装置の処理動作を説明するフローチャートであり、以下、これについて説明する。

【0030】(1) ストレージサブシステム12の始動時、ディスクアレ制御装置13は、自装置13と接続されている物理記憶装置15に対して、物理記憶装置15内の物理記憶領域23にある論理記憶領域のアドレスと実際にその論理記憶領域が存在する物理記憶領域のアドレスとの対応付け情報である論理/物理対応情報141、クラス構成情報142、クラス属性情報143等のディスクアレ構成情報14を送信する（ステップ300、310）。

【0031】(2) 次に、ディスクアレ制御装置13は、前述した情報の送信により、物理記憶装置15がアクセス可能となったときに、物理記憶領域制御装置22から送られてくる物理記憶装置15がアクセス可能なレディ状態に遷移した通知を受信する。このとき、物理記憶装置15は、ディスクアレ構成情報14による初期化終了の状態となっている（ステップ320）。

【0032】(3) 続いて、ディスクアレ制御装置13は、I/Oバス18経由でホスト10よりストレージサブシステム12に、そのストレージサブシステム12内の論理記憶領域に対してリードやライト処理のホストI/Oを送信してきたものや、ディスクアレ制御装置同士で命令やデータを受け渡すもの等の様々なデータを受信する（ステップ330）。

【0033】(4) 前記受信データとしてホストI/Oを受信した場合、ディスクアレ制御装置13は、ホストI/Oにより指定された論理記憶領域に対するリードまたはライト要求を受信し、その論理記憶領域のアドレス（論理アドレス）を物理記憶領域のアドレス（物理アドレス）に変換する論理/物理対応情報141を用いて、その論理記憶領域アドレスと対応する物理記憶領域23のアドレスを求める（ステップ340、350）。

【0034】(5) ディスクアレ制御装置13は、所定のデータが存在する物理記憶領域のアドレスを指定し、リード処理の場合、前述の物理アドレスを有する物理記憶装置からリードデータを読み出し、ホスト10にリードデータを転送し、ライト処理の場合、ホスト10から転送されたライトデータを受信し、その物理アドレスを持つ物理記憶装置にライトデータを転送する（360）。

【0035】図4は論理記憶領域のアドレスと物理記憶

領域のアドレスの対応が変化した場合のディスクアレイ制御装置の処理動作を説明するフローチャートであり、以下、これについて説明する。

【0036】(1) ディスクアレイ制御装置13は、物理記憶装置15の増減やRAIDレベルの変化、論理記憶領域が現在ある物理記憶領域アドレスとは別の物理記憶領域のアドレスに移動する等によって、論理記憶領域のアドレスと物理記憶領域のアドレスの対応が変化したことを監視し、変化があった場合、再度、物理記憶装置15に対して、物理記憶装置15内の物理記憶領域23にある論理記憶領域のアドレスと実際にその論理記憶領域が存在する物理記憶領域のアドレスとの対応付け情報である論理／物理対応情報141、クラス構成情報142、クラス属性情報143等のディスクアレイ構成情報14を送信する(ステップ3101)。

【0037】(2) 次に、ディスクアレイ制御装置13は、前述した情報の送信により、物理記憶装置15がアクセス可能となったときに、物理記憶領域制御装置22から送られてくる物理記憶装置15がアクセス可能なレディ状態に遷移した通知を受信する。このとき、物理記憶装置15は、ディスクアレイ構成情報14による更新終了の状態となっている(ステップ3201)。

【0038】(3) その後の処理は、図3により説明したステップ3300、3400、3500、3600の場合と同様に実行される(ステップ3301、3401、3501、3601)。

【0039】なお、前述したステップ320、3201において、ディスクアレイ制御装置13は、物理記憶領域制御装置22から、物理記憶装置15がアクセス可能状態に遷移したという情報を受信しなくてもよい。この場合、ある定まった時間後に物理記憶装置15に対してリードやライト等のアクセス可能な状態になっていると仮定し、何らかの記憶領域へのアクセス指示がディスクアレイ制御装置13にきた場合に、物理記憶装置15に所定の処理を行うためのアクセス処理を行えばよい。また、所定のアクセス処理に対する応答がない場合、再度、所定の処理を行うためアクセス処理を行うか、あるいは、応答が帰ってくるまで待ち、一定時間中に応答がない場合、何らかの記憶領域へのアクセス指示を出したモジュールに対してその旨を伝える方式としてもよい。

【0040】また、前述における論理／物理対応情報141は、論理記憶領域と物理記憶領域とを対応させる情報である。そして、論理アドレスは、ホスト10が前記リード／ライト処理部130で用いる論理記憶領域を示すアドレスである。また、物理アドレスは、実際にデータが格納される物理記憶装置15上の領域を示すアドレスであり、物理記憶装置番号及び物理記憶装置内アドレスからなる。記憶装置番号は、個々の物理記憶装置15を示す。記憶装置内アドレスは、物理記憶装置15内での記憶領域を示すアドレスである。

【0041】図5はディスクアレイ制御装置13が物理記憶装置15の稼動情報取得部223内の情報を読み出す際のディスクアレイ制御装置13の処理動作を説明するフローチャートであり、以下、これについて説明する。

【0042】(1) ディスクアレイ制御装置13は、ストレージサブシステム12が起動された後、記憶装置占有時間情報14Aを初期化し、その後、接続されている複数の物理記憶装置15にその物理記憶装置15のアクセス占有時間情報の取得要求を送信する(ステップ371、372)。

【0043】(2) 次に、ディスクアレイ制御装置13は、各物理記憶装置15よりアクセス占有時間情報を受け取り、各物理記憶装置のアクセス占有時間情報を記憶装置占有時間情報14Aに格納する(ステップ373、374)。

【0044】なお、前述したディスクアレイ制御装置13のアクセス占有情報の取得のタイミングは、ホストI/Oやバックアップ等によるその他のモジュールから物理記憶装置23に対するアクセスによるアクセス占有情報を各物理記憶装置15内の占有時間集計部227から一定時間間隔に読み出す方式や、他のモジュール(例えばホスト10や制御端末17)からアクセス占有時間情報取得要求がディスクアレイ制御装置13に送信された際等、様々であり、設計に依存する。

【0045】前述により取得されたアクセス占有時間情報は、ディスクアレイ制御装置13内の占有時間集計テーブルに記録される。

【0046】図6は物理記憶装置15内の物理記憶領域制御装置22の処理動作を説明するフローチャートであり、次に、これについて説明する。

【0047】(1) 物理記憶領域制御装置22は、ストレージサブシステム12の始動時にディスクアレイ制御装置13より送信されるデータ、すなわち、物理記憶装置15の物理記憶領域23にある論理アドレスと物理アドレスとの対応等の情報であるディスクアレイ構成情報14を受信する(ステップ400、401)。

【0048】(2) ディスクアレイ構成情報14を受信した物理記憶領域制御装置22は、その情報を元に、稼動情報取得部223内の構成・分類情報管理部226の構成・分類情報管理テーブルや占有時間集計部227の占有時間集計テーブルの作成初期化を行う(ステップ402)。

【0049】(3) ステップ402でのテーブルの初期化処理が終了すると、この物理記憶装置15にアクセス可能であることを認識させるためディスクアレイ制御装置13に初期化処理が終了したことを通知する。なお、ディスクアレイ制御装置13に物理ディスク装置がアクセス可能な状態に遷移した情報を送信しなくてもよい。この場合、定まった時間後に物理記憶装置15の物理記

憶領域制御装置22に対して何らかの記憶領域へのアクセス指示がきた場合、物理記憶領域制御装置22は、所定の処理を行うために物理記憶領域23にアクセス可能な状態であれば、その物理記憶領域23にアクセスし所定の処理を行い、不可能であれば、アクセス要求情報記憶部225にアクセス情報を格納し、物理記憶領域にアクセス可能になった状態で所定の処理を行うか、あるいは、所定の処理を行うために物理記憶領域23にアクセス可能な状態になるまで、記憶領域への何らかのアクセス指示を受け付けないようにしてもよい(ステップ403)。

【0050】(4)その後、物理記憶領域制御装置22は、ディスクアレ制御装置13からホストI/Oや、物理記憶装置移動情報取得要求命令、あるいは、新たなディスクアレ構成情報が送信されてくるのを待ってそれを受信する(ステップ404)。

【0051】(5)ステップ404で、ディスクアレ制御装置13からホストI/Oを受信すると、そのI/Oがリード処理かライト処理かを判定し、リード処理であった場合、データ処理制御部224は、読み出すべきデータがキャッシュメモリ22内に存在するか否かをチェックする(ステップ405、406)。

【0052】(6)ステップ406のチェックで、そのデータがキャッシュメモリ22内に存在した場合、そのデータをキャッシュメモリ22から読み出し、また、そのデータがキャッシュメモリ22内に存在しなかった場合、物理記憶領域23からそのデータを読み出して、データをディスクアレ制御装置13に転送する(ステップ407、709、408)。

【0053】(7)ステップ405で、ホストI/Oがライト処理であると判定された場合、データ処理制御部224は、ホスト10から転送されたライトデータを受信し、キャッシュメモリ22にそのライトデータを書き込む(ステップ410、411)。

【0054】(8)そして、データ処理制御部224は、データ書き込み終了通知をディスクアレ制御装置13に通知すると共に、前述のライトデータを物理記憶領域23に格納する(ステップ412、413)。

【0055】(9)ステップ408の処理後、または、ステップ413の処理後、データ処理制御部224は、キャッシュメモリ22にアクセスしたか、あるいは、物理記憶領域23にアクセスしたかの情報、ランダムリードかシーケンシャルリードか等のJOB種別情報、ライトデータを物理記憶領域23に書き込む際のランダムリード/ライトかシーケンシャルリード/ライトか等のJOB種別情報のアクセス種別を認識し、アクセス種別毎に、キャッシュメモリ22あるいは物理記憶領域23にアクセスした占有時間情報を移動情報取得部223内の占有時間集計部227に格納する(ステップ414、415)。

【0056】(10)ステップ404でディスクアレ制御装置13から新たなディスクアレ構成情報を受信すると、データ処理制御部224は、移動情報取得部223の構成・分類情報管理部226内の情報を新たなディスクアレ構成情報に対応して書き換える(ステップ418、419)。

【0057】(11)ステップ404でディスクアレ制御装置13から物理記憶装置移動情報取得要求命令を受信した場合、データ処理制御部224は、移動情報取得部223内の占有時間集計部227に格納している物理記憶装置15のアクセス占有時間情報を読み出して、それをディスクアレ制御装置13に送信する(ステップ416、417)。

【0058】なお、前述したステップ417の処理での占有時間情報の物理記憶装置15からディスクアレ制御装置13への送信は、一定時間間隔で物理記憶装置15からディスクアレ制御装置13に自動的に行うようにしてもよい。この場合、ディスクアレ制御装置13から前記物理記憶装置移動情報取得要求命令が物理記憶装置13に送信されてくることはない。

【0059】図7はディスクアレ制御装置13内に保持されている論理記憶領域のアドレスと物理記憶領域のアドレスとの対応を管理するためのテーブル内の論理/物理対応情報141の構成例を説明する図である。

【0060】ディスクアレ制御装置13は、接続されている複数の物理記憶装置15内の物理記憶領域23にある論理記憶領域のアドレスとその論理記憶領域内の物理記憶領域のアドレスとの対応を管理している。これに使用する論理/物理対応情報141は、図7に示すように、特定の論理記憶領域に対して付与される論理記憶領域番号500、論理アドレス510、その論理記憶領域がある物理的な記憶領域を持つ記憶装置番号521と物理的な記憶領域のアドレス522とによる物理アドレス520、その物理記憶装置15の性能を示すレイドレベル530、その物理記憶装置15が属しているパリティグループ番号540のそれぞれが対応付けられて構成される。ディスクアレ制御装置13は、このような論理/物理対応情報141を有することにより、ホスト10や制御端末17、その他のモジュール(例えば、その他のディスクアレ制御装置等)からリードやライト等の処理が論理記憶領域のアドレスを指定して自ディスクアレ制御装置13に対してアクセスされた場合、論理領域のアドレスを物理的な記憶領域のアドレスに変換して、データのリード/ライト処理を物理記憶装置15に対して正確に行うことができる。

【0061】図8はディスクアレ制御装置13内に格納される論理領域使用状況情報144と物理領域使用状況情報145等の記憶装置占有時間情報141の例を示す図である。これらの情報は、占有時間集計テーブルとして構成されている。

【0062】ディスクアレイ制御装置13は、ホストI/Oやバックアップ等によるその他のモジュールからのアクセスによる物理記憶装置15の物理記憶領域23へのアクセス占有情報を各物理記憶装置15内の占有時間集計部227から定期的に読み出して、そのアクセス占有時間情報を受信ディスクアレイ制御装置13内の占有時間集計テーブルに記録する。図8に示す例は、論理記憶領域番号601毎、I/O JOB種別602毎に占有時間を集計したものである。I/O JOB種別として、図示例では、シーケンシャルリード610、シーケンシャルライトデータ620、シーケンシャルライトパリティ630、ランダムリード640、ランダムライトパリティ660、キャッシュヒット時670、合計680が示されているが、さらに他のI/O JOB種別があってもよい。

【0063】物理記憶装置15から読み出される占有時間は、前述に限らず、I/O毎のアクセス占有時間の累積値や、ユニバーサルな時間値、マシン固有な時間値による記録であってもよい。また、ディスクアレイ制御装置13において、各論理記憶領域や物理記憶領域のアクセス占有時間を編集して、物理記憶装置毎やパリティグループ毎のアクセス占有時間を求めた値に基づいて新たに占有時間情報テーブルを作成してもよい。

【0064】前述により、ホスト10や制御端末17等が物理記憶装置の移動情報取得要求をディスクアレイサブシステム12に出した場合、ホスト10や制御端末17等は、直接物理記憶装置15にアクセスして論理記憶領域や物理記憶領域のアクセス占有時間情報を取得しなくてもディスクアレイ制御装置13からその論理記憶領域や物理記憶領域のアクセス占有時間情報を取得することが可能となる。

【0065】図9は物理記憶装置15内の移動情報取得部223内の構成・分類情報管理部226に格納される論理記憶領域のアドレスと物理記憶領域のアドレスとの対応付けを管理するテーブルの構成例を示す図である。

【0066】物理記憶装置15内のデータ処理制御部224は、ディスクアレイ制御装置13より、ストレージサブシステム12の始動時や、物理記憶装置の増減やレイドレベルの変化、論理記憶領域の移動等によって生じる論理記憶領域のアドレスと物理記憶領域のアドレスとの関係が変化した際に、その論理記憶領域と物理記憶領域との対応付け情報を受信し、それを構成・分類情報管理部226に図9に示すような対応テーブルに格納する。この対応テーブルは、特定の論理記憶領域に対して付与される論理記憶領域番号700、論理アドレス710、その論理記憶領域がある物理的な記憶領域を持つ記憶装置番号721と物理的な記憶領域のアドレス722とによる物理アドレス7520により構成される。これにより、物理記憶装置15は、どのアドレス論理記憶領域が自物理記憶装置15内の物理記憶領域アドレスのどこ

にあるかを認識することができる。

【0067】図10は物理記憶装置15内の移動情報取得部223の占有時間集計部227に累積・格納されるI/Oによる記憶領域へのアクセスによる占有時間情報のテーブルの構成例を示す図である。

【0068】このテーブルは、論理記憶領域番号801毎、I/O JOB種別802毎に占有時間を集計したものである。I/O JOB種別として、図示例では、シーケンシャルリード810、シーケンシャルライトデータ820、シーケンシャルライトパリティ830、ランダムリード840、ランダムライトパリティ860、キャッシュヒット時870、合計880が示されているが、さらに他のI/O JOB種別があってもよい。

【0069】物理記憶装置15内のデータ処理制御部224は、ホストからのI/O等により記憶装置にアクセスがあると、アクセスがあった各論理記憶領域801に関して、アクセスのJOB種別802毎に、占有時間890を占有時間集計部227に累積する。これにより、物理記憶装置15内の論理記憶領域の番号と、その論理記憶領域へのアクセス種別による占有時間と、ある時間内の合計占有時間との関係を物理記憶装置15内で得ることが可能となる。

【0070】前述した本発明の実施形態によれば、I/O毎の記憶装置の占有時間情報を取得することができ、かつ、記憶領域へのアクセスによる記憶装置の占有時間情報の取得を物理記憶装置内で実現することができる。

【0071】本発明の実施形態によれば、前述により、ストレージサブシステムを構成する複数の物理記憶装置のそれぞれが、物理記憶装置の論理記憶領域の占有時間を取得することが可能となり、ディスクアレイ制御装置のみでは取得不可能なシステム構成で各論理記憶領域の占有時間（実稼働時間）取得することが可能となる。また、前述した本発明の実施形態によれば、1つのI/O毎の物理記憶装置への影響を考慮することによって、前記ストレージサブシステムの論理記憶領域の利用率の解析や利用率予測の誤差を小さくすることができ、より最適な性能チューニングを行うために、より精度の高い物理記憶装置へのI/O毎のアクセス占有時間情報の取得を行うことができる。

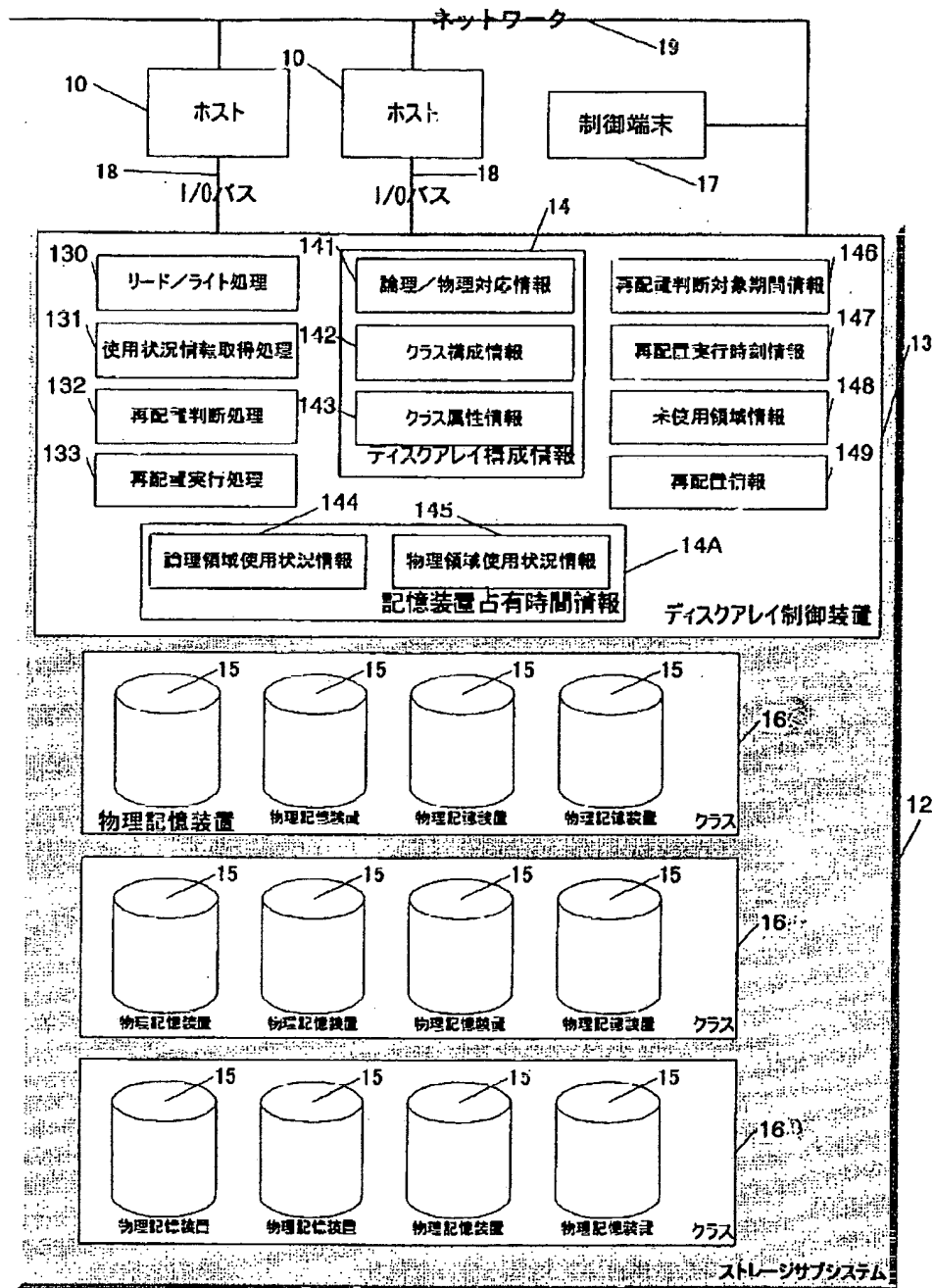
【0072】

【発明の効果】以上説明したように本発明によれば、物理記憶装置が論理記憶領域の占有時間を取得することができ、ディスクアレイ制御装置のみでは取得不可能なシステム構成で各論理記憶領域の占有時間（実稼働時間）を取得することができる。

【0073】また、本発明によれば、物理記憶装置へのI/O毎のアクセス占有時間情報を取得することができるため、1つのI/O毎の物理記憶装置への影響を考慮したストレージサブシステムの論理記憶領域の利用率の解析や利用率の予測を小さい誤差で行うことが可能とな

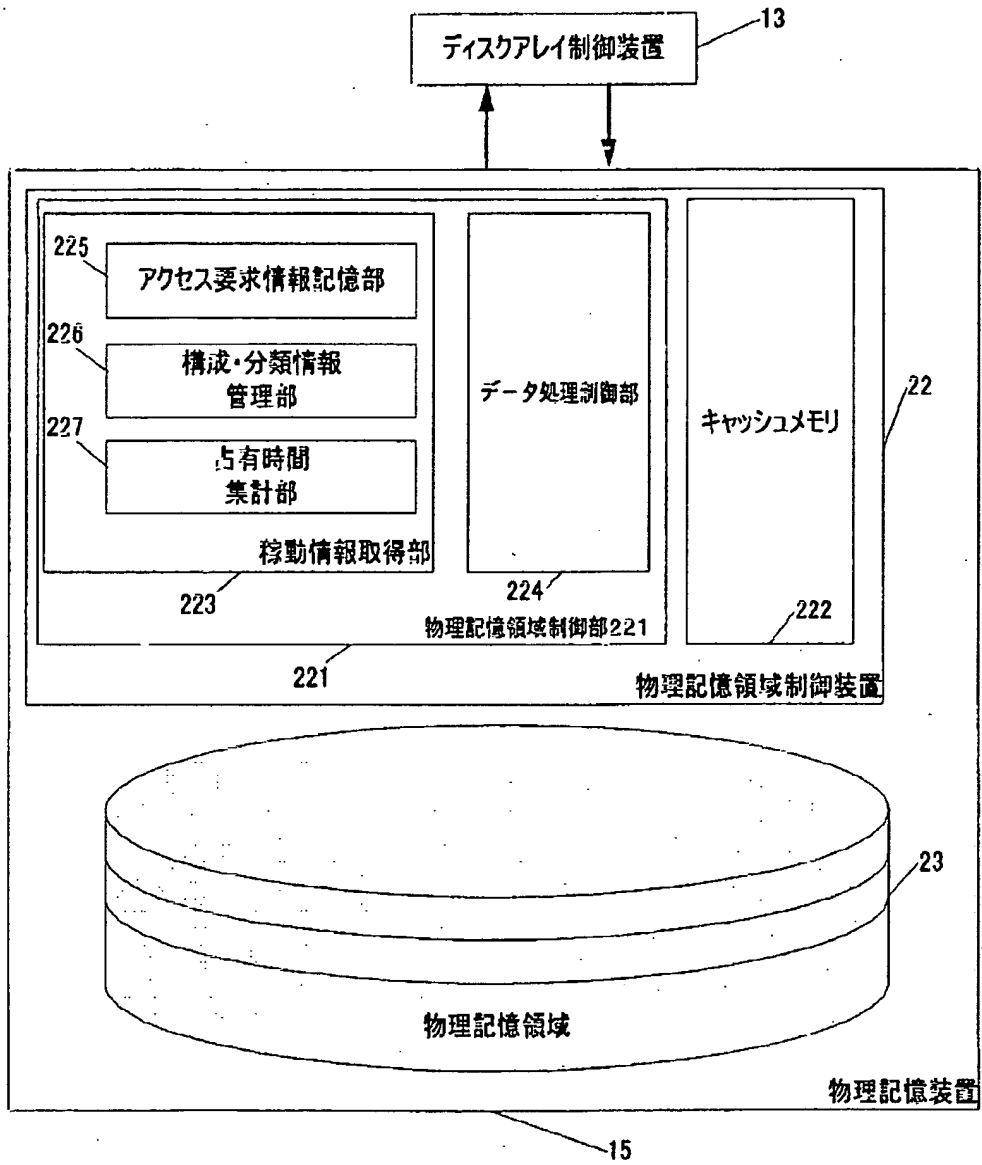
【図1】

図1



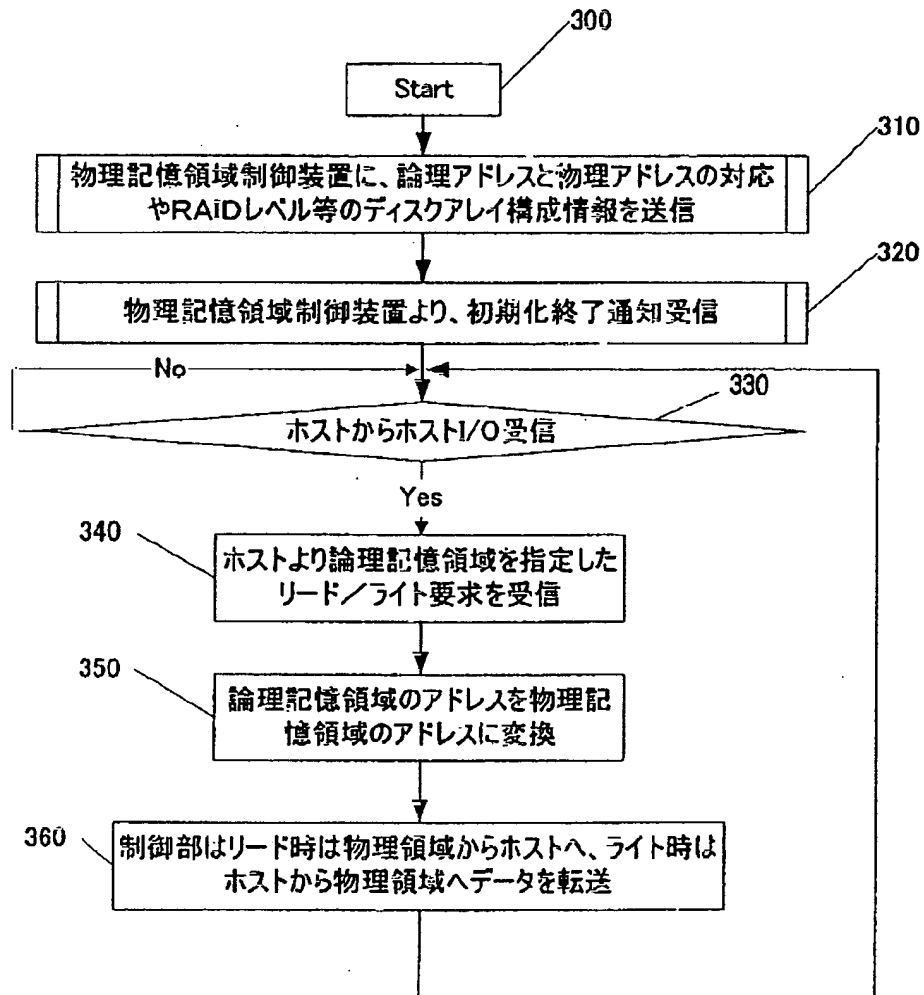
【図2】

図2



【図3】

図3



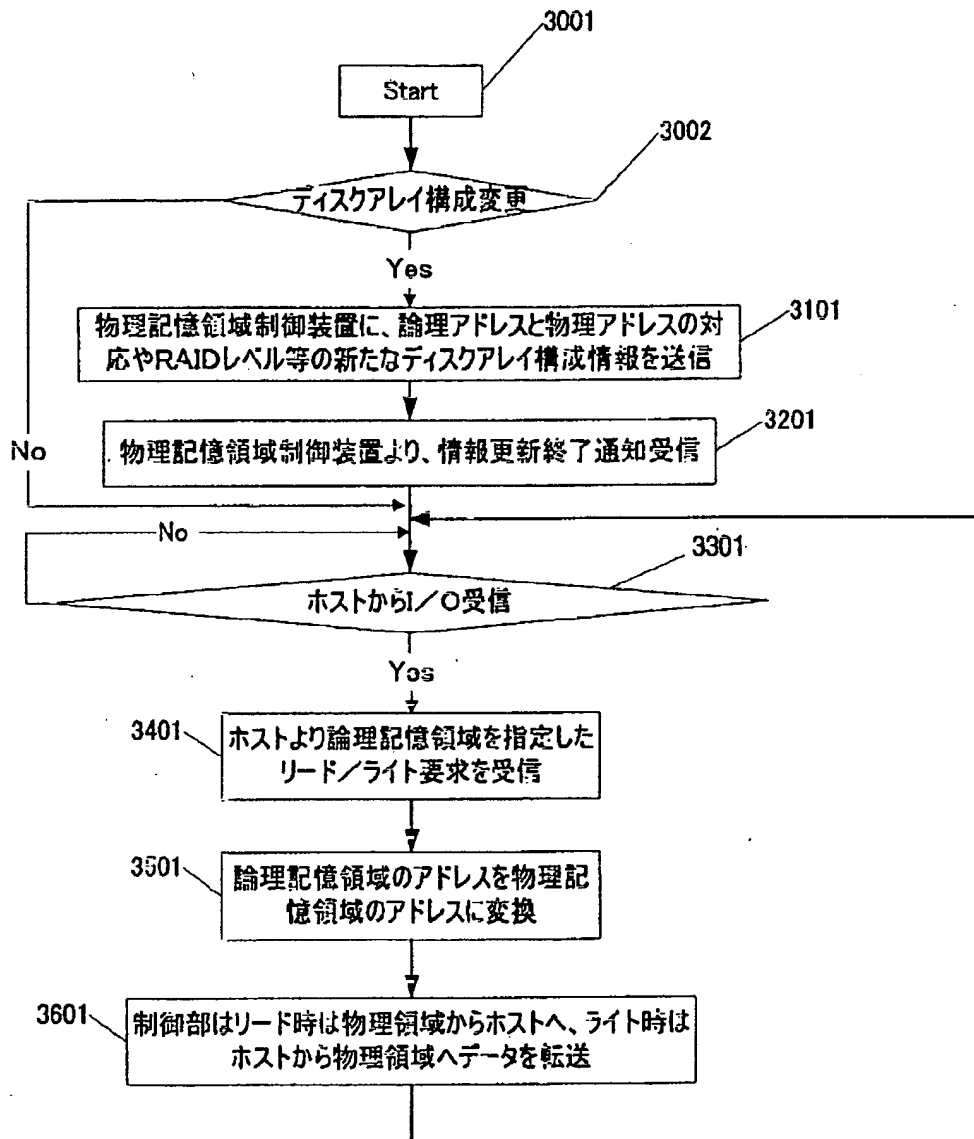
【図9】

図9

論理記憶 領域番号	論理アドレス	物理アドレス	
		物理記憶装置番号	物理記憶領域アドレス
3	3000~3999	1	0~999
4	4000~4999	1	1000~1999
5	5000~5999	1	2000~2999

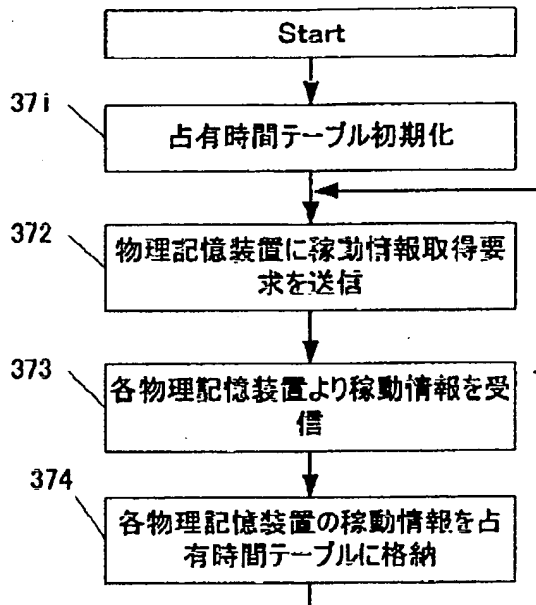
【図4】

図 4



【図5】

図5



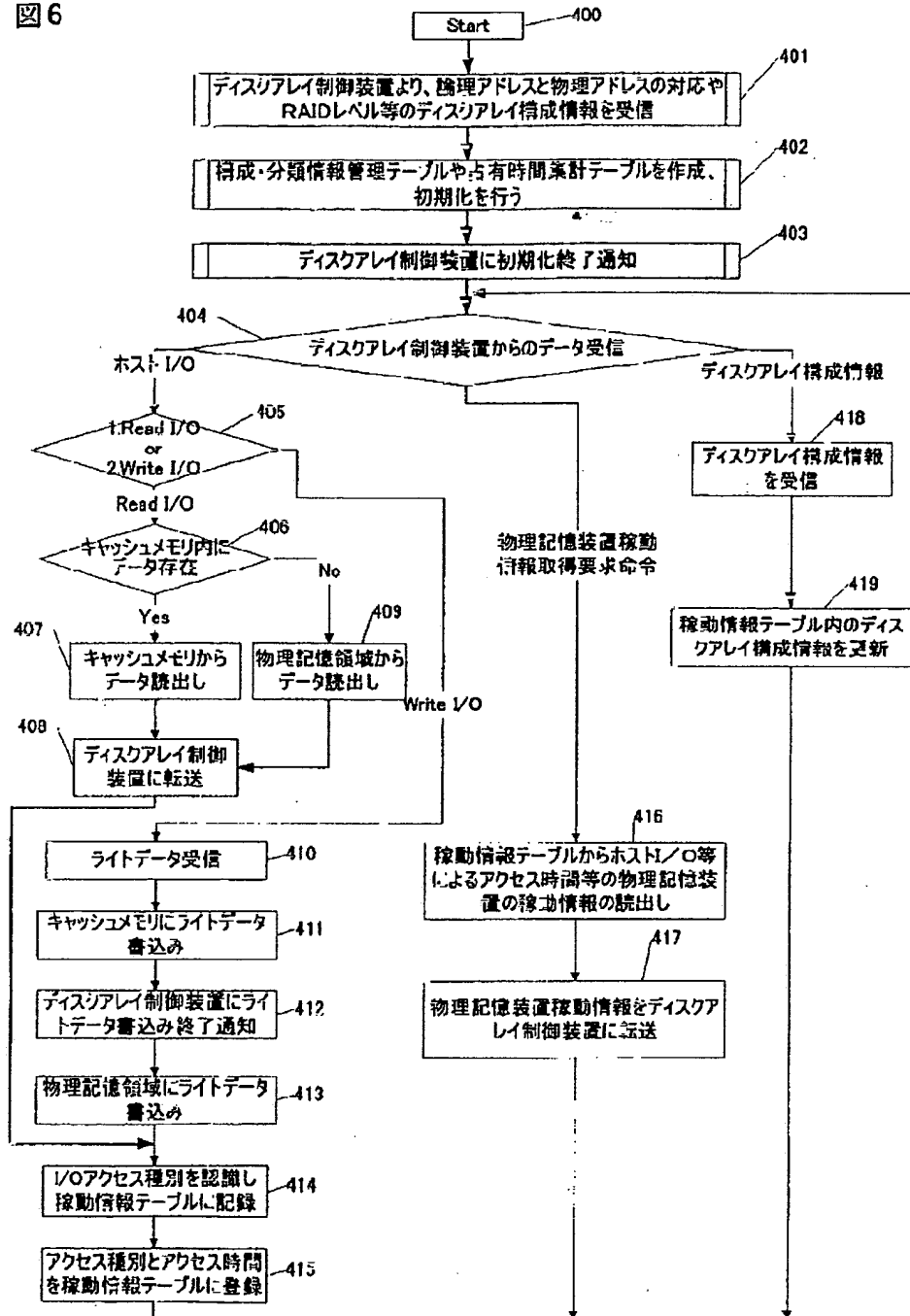
【図7】

図7

論理記憶 領域番号	論理アドレス	物理アドレス		レイドレベル	パリティグ ループ番号
		物理記憶装置番号	物理記憶領域アドレス		
0	0~999	0	0~999	1	100
1	1000~1999	0	1000~1999	1	100
2	2000~2999	0	2000~2999	1	100
3	3000~3999	1	0~999	5	120
4	4000~4999	1	1000~1999	5	120
5	5000~5999	1	2000~2999	5	120
...

【図6】

図 6



【図10】

図 10

801 論理記憶領域 番号	802 I/O JOB種 別	810 Sequential Read	820 Sequential Write data	830 Sequential Write Parity	840 Random Read	850 Random write Data	860 Random write Parity	870 キャッシュ ヒット時	880 合計
3		占有時間	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間
4		占有時間	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間
5		占有時間	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間

890

 フロントページの続き

(72)発明者 荒川 敬史

 神奈川県川崎市麻生区王禅寺1099番地 株
 式会社日立製作所システム開発研究所内

(72)発明者 大枝 高

 神奈川県川崎市麻生区王禅寺1099番地 株
 式会社日立製作所システム開発研究所内

(72)発明者 荒井 弘治

 神奈川県小田原市国府津2880番地 株式会
 社日立製作所ストレージシステム事業部内

F ターム(参考) 5B018 GA07 MA14 QA16

5B065 BA01 BA06 CA13 CA30 CC03

CH19 ZA02

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.